ORGANIC LETTERS

2004
Vol. 6, No. 11
1825−1827

What Are the Limits to the Size of Effective Dynamic Combinatorial Libraries?

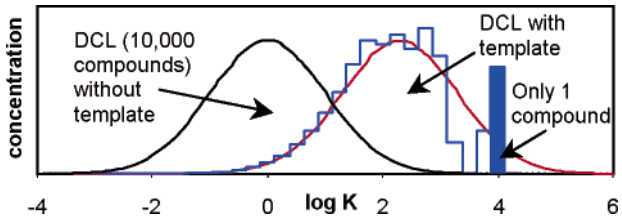
Peter T. Corbett, Sijbren Otto,* and Jeremy K. M. Sanders*

Cambridge University Chemical Laboratory, Lensfield Road, Cambridge CB2 1EW, U.K.

so230@cam.ac.uk; jkms@cam.ac.uk

Received April 1, 2004

ABSTRACT



Using simple computer simulations of model dynamic combinatorial libraries, we show that the best binders can be amplified to useful concentrations in libraries containing 10–10^6 compounds.

Dynamic combinatorial chemistry is a powerful new approach to the discovery of compounds that strongly engage in molecular recognition.[1] A dynamic combinatorial library (DCL) is created from a set of building blocks, connected using reversible reactions to generate a mixture at thermodynamic equilibrium. Reversible host−guest binding to an added template shifts the equilibrium toward library members that bind to it, resulting in the selective amplification of strong binders. In practice, a typical DCL experiment consists of comparing the concentrations of all compounds in the library in the absence of template with the corresponding concentrations in the presence of template. Any compound that has increased in concentration should be a good binder.

A variety of DCL experiments have shown significant amplification of strong binders using a range of architectures and reversible reactions. However, in most of these DCLs only relatively low levels of diversity have been sampled.[2] Whereas these libraries have already produced many new successful compounds, a question remains as to the degree of diversity that can practically be sampled using this technique: i.e., how many compounds can a DCL contain while still allowing the amplification of strong binders to useful concentrations?

The only published theoretical study that addresses amplification in large libraries considers the limiting case of an infinitely large DCL.[3] The effect of amplification upon a continuum of library members with a log-normal distribution of binding constants was considered: with a large excess of template, it was shown that the mean[4] binding constant in the DCL was shifted upward by slightly more than 2 orders of magnitude.

While such a model provides powerful insight into the behavior of the bulk of the library members in a diverse DCL,

(1) Reviews: (a) Otto, S. *Curr. Opin. Drug Discovery Dev.* **2003**, *6*, 509−520. (b) Otto, S.; Furlan, R. L. E.; Sanders, J. K. M. *Curr. Opin. Chem. Biol.* **2002**, *6*, 295−321. (c) Rowan, S. J.; Cantrill, S. J.; Cousins, G. R. L.; Sanders, J. K. M.; Stoddart, J. F. *Angew. Chem., Int. Ed.* **2002**, *41*, 898−952. (d) Ramström, O.; Bunyapaboonsri, T.; Lohmann, S.; Lehn, J.-M. *Biochim. Biophys. Acta* **2002**, *1572*, 178−186. (e) Otto, S.; Furlan, R. L. E.; Sanders, J. K. M. *Drug Discovery Today* **2002**, *7*, 117−125.

(2) Some systems—for example, DCLs of macrocycles, where there is no clearly defined limit to the size of the macrocycles—contain a very large number of compounds. However, the majority of those compounds (e.g., very large macrocycles) are likely to be present at such low concentrations in the absence of template (to the extent that all of these compounds together may only account for a small fraction of the library) that even if they were to be strongly amplified, this would go unnoticed. Detectable amplification is likely to be confined to the relatively small set of compounds with appreciable concentrations.

(3) Moore, J. S.; Zimmerman, N. W. *Org. Lett.* **2000**, *2*, 915−918.

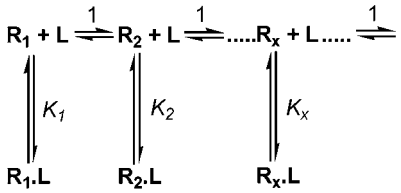
(4) The *geometric* mean, as the relevant distributions are log-normal.

10.1021/ol049398k CCC: $27.50
Published on Web 04/27/2004

it does not directly predict the concentration of the *single best library member* in a large but finite DCL. Yet when using DCLs for the discovery of strong binders the real interest is in the one or few outliers rather than in the bulk of the library. Herein, we extend the model to allow an assessment of the changes in concentration of the best binder upon introduction of a template. The simplest way to do this is to simulate equilibria involving a large set of discrete library members, each with a randomly chosen binding constant.

Simulated DCLs were constructed using the same assumptions as used by Moore and Zimmerman: a collection of host molecules[5] is considered that can reversibly interconvert (Scheme 1). All hosts are present at the same initial concentration. Each host is then randomly assigned a binding constant for the template, such that the values of log *K* give a normal distribution, with a mean[6] of 0 and a standard deviation of 1. A large excess of template is used (10 M template to 1 M total host concentration) to create a situation in which almost all library members are bound to the template.

Scheme 1. Simple Model of a DCL



These simple assumptions constitute an idealized DCL. Every host in the library can interconvert into any other, and in the absence of guest, all hosts are present at the same initial concentration. In practice, DCLs often contain a number of different building blocks, and/or different numbers of building blocks per host are permitted. This leads to variation in the concentrations of the various hosts in the absence of template, which will bias the corresponding concentrations in the presence of the template. Furthermore, it has been shown that, depending on the experimental conditions, the competition of different hosts for the different building blocks can create a partial breakdown in the correlation between amplification and binding affinity.[7,8] Our present model does not consider these effects, thus allowing a separate analysis of the effect of library size on amplification.

We have previously developed a computer program (DCLSim[8]) that calculates the concentration of compounds in large equilibrium mixtures in the absence and presence of a template. This program was used to calculate the concentration of all compounds in a 10 000 membered DCL in the presence of a template. Each of the library members was randomly assigned a binding constant on the basis of a log-normal distribution. Figure 1 shows how the hosts in the library are distributed over various affinity classes (i.e., ranges of log *K*) and how many compounds each range contains. The solid lines in Figure 1 represent the unshifted (no template) and shifted (template present) distributions as published by Moore.[3] For low binding constants, the simulated distribution matches Moore's continuum distribution very well. However, at the extreme of high binding, the tail end of the shifted distribution is concentrated into a single compound that in this example is amplified by a factor of 794 to become 8% of the library material.

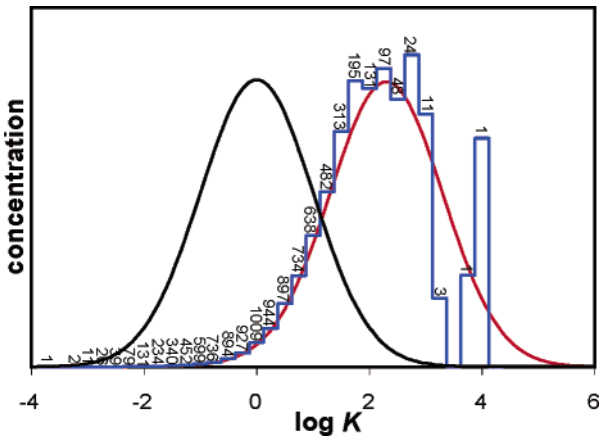


Figure 1. Histograms representing the composition of a continuous DCL in the absence (black) and presence (red) of a template[3] and a typical simulated DCL containing 10 000 compounds (blue). Bars are labeled with the number of compounds in the affinity class.

The results in Figure 1 represent only one example for one library size and one particular distribution of binding constants. To explore the relationship between amplification and library size in more detail, we have simulated libraries ranging in size from 10 compounds to a million.[9] Since the binding constants are assigned to the hosts randomly, the extent of amplification of the best binders will vary from one simulation to the next. To obtain statistically significant data, we have carried out 100 simulations for each library size and calculated the mean yield of the best binder (Figure 2).[10] As expected, the yield of the best binder declines with

(5) All of the models discussed in this paper apply equally well to DCLs where the library members act as the guests, and the template as a host.

(6) As with Moore's model,[3] variation of this value does not significantly affect the main conclusions of this paper. Larger values of this mean will simply result in an increase in the mean and highest binding constants in the templated library. In experimental systems, such an elevated mean could arise from the use of a carefully chosen set of building blocks, with features that are likely to be complementary to the template.

(7) (a) Grote, Z.; Scopelliti, R.; Severin, K. *Angew. Chem., Int. Ed.* **2003**, *42*, 3821−3825. (b) Severin, K. *Chem. Eur. J.*, in press.

(8) Corbett, P. T.; Otto, S.; Sanders, J. K. M. *Chem. Eur. J.*, in press.

(9) For DCLs with 10 000 or more compounds, explicitly simulating each of the library members individually became too computationally demanding, so an approximation was introduced. For each of these DCLs, a threshold log *K* was chosen. Library members with binding constants above the threshold (approximately 1 in 100) were all included in the simulation. Only one in *n* of those with binding constants below the threshold were included, but the equilibrium constants for the formation of those library members were set so as to increase their concentration in the template-free library by a factor of *n*. In all of these cases, *n* was chosen such that the sub-threshold part of the library was represented by 100 compounds.

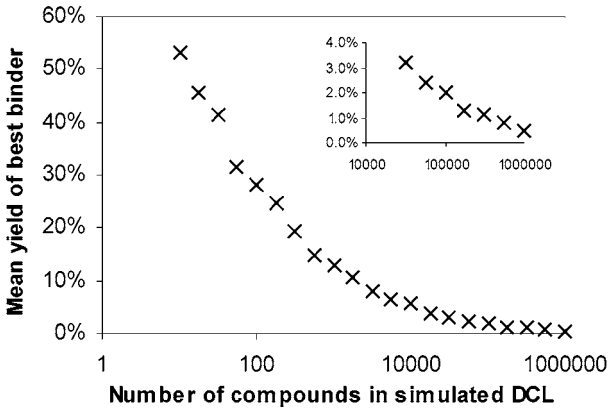


Figure 2. Geometric mean yields of the most highly amplified hosts from 100 simulated DCLs per data point.

library size, but only relatively slowly. A 1000 compound library, for example, will allow for the best compound to become, on average, 13% of the library. Even in extremely large libraries the best compound can be amplified to usable fractions—in a 100 000 compound library the best binder accounts for 2% of the total material, whereas in a 1 million compound library this fraction is 0.5%. For large libraries, the concentration of the best compound appears to be roughly inversely proportional to the square root of the library size.

This relatively slow decay in the concentration of the best compound upon increasing the library size can be explained as the result of two competing effects. As the size of the library increases, the chance that it contains an exceptionally good binder increases; i.e., the best binder in a large library will generally be better than the best binder in a smaller library. This effect is illustrated in Figure 3a. In general, higher binding constants lead to greater amplifications, and so in larger libraries the best binders will be amplified to a greater extent (Figure 3b). However, a larger set of compounds implies a lower initial concentration for all of the library members, which in turn implies that the amplified concentrations will be lower. With large numbers of compounds, the latter effect appears to dominate.

Although the quantities of the best binder produced from very diverse DCLs are far below the yields that might be expected from conventional synthetic reactions, it should be noted that in many situations, high yields are not required at the screening stage. It is merely necessary to be able to identify the composition of the most highly amplified compound. With this information, a *biased library*[11] can be constructed using only the required building blocks in the correct proportions, which should produce the identified receptor in high yield.

(10) An alternative approach is to consider the expected binding affinity of the best compound in the DCL, and take the area of the "tail" of the graph for the shifted equilibrium above that value to be the concentration of the best binder. This method generally overestimates concentrations of the top compounds—but never by more than a factor of 2.

(11) Otto, S.; Furlan, R. L. E.; Sanders, J. K. M. *Science* **2002**, *297*, 590−593.

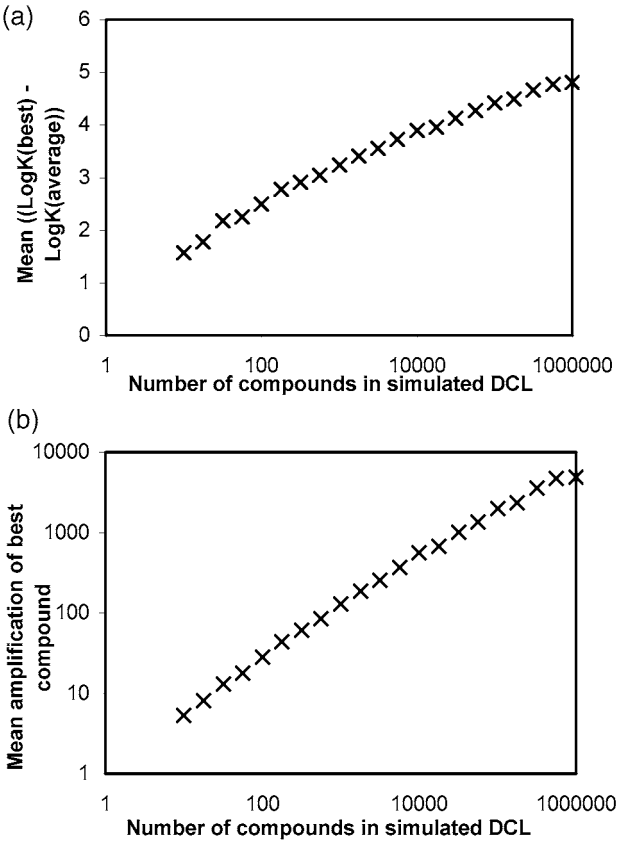


Figure 3. (a) Increase in binding constant above the mean of the most highly amplified hosts from 100 simulated DCLs per data point. (b) Mean amplification factors of the best compounds ($[best]_{templated}/[best]_{untemplated}$) in 100 simulated DCLs per data point.

In conclusion, our simulations indicate that, even in model DCLs containing up to 10^6 compounds, host−guest binding can induce the amplification of strong binders to concentrations that are easily within the detection limits of modern analytical equipment. Studies are currently underway to simulate libraries where hosts can contain different building blocks and can have unequal starting concentrations in order to assess (1) how these factors influence amplification relative to the "benchmark" established by this paper and (2) what experimental conditions and architectures can be used to bring the results closer to this ideal.

Acknowledgment. We acknowledge support from the Royal Society to S.O. and from EPSRC to J.K.M.S. and P.T.C..

Supporting Information Available: A spreadsheet containing the details of the best binders in the libraries used to generate Figures 2 and 3. This material is available free of charge via the Internet at http://pubs.acs.org.

OL049398K